**Question 1: Concept Review**

1. Training MSE will decrease from 0 (exponential curve). Test MSE will follow a U shape towards the right side of training.

A diagram of error and test error

Description automatically generated with medium confidence

1. Perfectly separated data. Logistic regression fails in this setting because the coefficients can’t converge because the maximum likelihood estimation is maximized to infinity

A graph on a white paper

Description automatically generated

1. False. You can’t directly calculate bias and variance of a regularized regression model for a given dataset to determine the trade-off between bias and variance and choose an optimal λ. Bias and variance analysis typically require knowledge of the true population. Bias and variance are statistical properties that describe how well a model generalizes to unseen data and how much it varies across different training sets. To choose an optimal λ, use techniques like cross-validation, while involve splitting the dataset into training and validation sets to estimate how well the model will perform on new, unseen data, without the need for knowledge of the true population.
2. False, this is not always true. The choice between Ridge and Lasso depends on the specific dataset and problem. Ridge tends to perform well when a large set of predictors are associated with the response variable, but it doesn’t enforce sparsity and will include all predictors to some extent. Lasso, on the other hand encourages sparsity by forcing some coefficients to be exactly zero, making it effective when only a subset of predictors is truly relevant . The relative performance of Ridge and Lasso depends on the data characteristics and the underlying relationships between predictors and the response variable.
3. True. QDA is equivalent to using Bayes Rule to approximate P(Y=k|X) under the assumption that the predictors follow a normal distribution. P(Y=1|X) = P(X|Y=1)P(Y=1) / P(X). QDA models the conditional probability of a class (Y) given the predictors (X) by calculating the likelihood of the data given the class, the prior probability of the class, and the marginal likelihood of the data. This approach aligns with the fundamental principles of Bayes Rule, where the posterior probability based on the likelihood and prior probability is calculated.
4. Yes, we can take the inverse but it wouldn’t be of full rank, and the classifier will get really large. If it is perfectly colinear, you can’t take the inverse

**Question 2: Simulations**

1. See alpha.fn R script
2. See alpha.fn R script